

Response to: OpenClaw and Multi-Agent AI in Plastic Surgery

Orr Shauly, MD^o

We thank the authors of the response to our manuscript for their thoughtful assessment of our evaluation of chatbot triage performance across plastic surgery websites. We mostly agree with their central premise. The single-model chatbot may be the wrong unit of intervention, and the natural next question is whether modular, task-specific agents within a multiagent framework can be matched to the discrete jobs that make up the entire perioperative communication environment. We welcome that reframing, and we welcome their proposal of a shared comparative system. We write to draw 2 boundaries around this line of work—1 clinical and 1 operational—that we believe are essential if it is to be pursued safely.

The first boundary is clinical and is one of the central tenants of our original discussion of this topic. No architecture removes the requirement for clinical judgment, and improved sensitivity does not change where final authority must sit. Our central finding was not that these systems were poorly engineered but that they failed to recognize the cases that most needed a clinician, missing most emergent presentations. That is a failure of recall at the reasoning layer, and decomposing 1 model into several agents built on the same underlying reasoning does not by itself create sensitivity that the model lacks. Even granting that multiagent designs eventually improve accuracy, the correct ceiling on autonomy for any urgent or emergent concern is triage and escalation, not disposition. An agent may widen the front door, and it may surface, summarize, and route more appropriately, but a clinician must remain the terminal node for anything that may require acute intervention. The hazard the authors themselves name, that an autonomous system can convert a wrong inference into a wrong action before any clinician sees it, is precisely the failure mode our data predict if escalation authority is pushed past the human. The lesson of an 80% missed-emergent rate is not to build a more confident autonomous triager but to guarantee that the human remains the last clinician in the chain.

The second boundary is operational, and here we must be direct, because the specific instrument the authors propose has no place in this discussion. OpenClaw is a single-user consumer personal-assistant application, built around 1 embedded agent that acts on an individual's behalf through consumer messaging apps.¹ It is not a multiagent orchestration framework, and grouping it with the tools used to build such systems amounts to a category error. Before any question of clinical or regulatory suitability arises, it is simply the

wrong kind of tool for the architecture the authors are advocating. Proposing it as the substrate on which to test a multiagent triage workflow is therefore not a debatable matter of tooling preference. It signals a fundamental misunderstanding of the systems under discussion.

The problems compound from there. Set aside, for a moment, that OpenClaw's community skill registry has been found by independent security analyses to contain a substantial fraction of skills exhibiting prompt injection, credential harvesting, or malware, and that its gateway binds to all network interfaces by default.² Even bracketing all of that, OpenClaw would not appear on any serious shortlist of frameworks for this task. Teams building multiagent clinical workflows reach for orchestration tooling such as LangGraph, CrewAI, or AutoGen, or for a managed agent service operated under a Business Associate Agreement such as Amazon Bedrock, and they deploy it inside a HIPAA (Health Insurance Portability and Accountability Act) -eligible environment with role-based access control, encryption, and audit logging, because compliance is a property of the deployment and the contracts, and never of the software alone.³ OpenClaw is absent from that landscape for reasons that have nothing to do with our particular use case and everything to do with what it is. That it was proposed at all, and proposed within the peer-reviewed record, is the most telling part of this exchange.

That is the deeper issue this exchange exposes—a 2-sided literacy gap that is itself a hazard to patients. Clinicians who do not understand these systems cannot govern them, and that is dangerous. Technologists who understand the systems but not the ethos of medicine—the duty of care, the privacy obligations, the asymmetric cost of a single missed emergency to the life of a human being—should not be the ones directing clinical implementation, and that is dangerous in a quieter way. The resolution is neither to keep artificial intelligence out of medicine nor to delegate medicine's judgment to those who build the tools. It is for clinicians to acquire genuine fluency and to lead implementation with technical expertise in support. Normalizing a basic

Dr Shauly a plastic surgery resident, Division of Plastic and Reconstructive Surgery, Emory University, Atlanta, GA, USA.

Corresponding Author:

Dr Orr Shauly, 3200 Downwood Cr NE, Atlanta, GA 30327, USA.
E-mail: oshauly@emory.edu; Instagram: @orshaulymd

Aesthetic Surgery Journal

2026, Vol 00(0) 1–2

Accepted date: July 10, 2026; online
publish-ahead-of-print July 24, 2026.

© The Author(s) 2026. Published by
Oxford University Press on behalf of
The Aesthetic Society. All rights
reserved. For commercial re-use,
please contact reprints@oup.com for
reprints and translation rights for
reprints. All other permissions can be
obtained through our RightsLink service
via the Permissions link on the article
page on our site—for further information
please contact
journals.permissions@oup.com.
<https://doi.org/10.1093/asj/sjag156>
www.aestheticsurgeryjournal.com

OXFORD
UNIVERSITY PRESS

consumer agent (OpenClaw) as an option for multiagent healthcare triage within the peer-reviewed record is exactly the kind of error this gap produces, and it lends unearned legitimacy to an unsafe pattern.

We therefore endorse the authors' call for prospective, comparative evaluation of monolithic and multiagent systems, with common case libraries, common escalation criteria, and common safety endpoints. We add 2 conditions—that such evaluation must run on infrastructure that already meets medicine's privacy and security obligations, and it must preserve the clinician as the final authority for every urgent and emergent disposition. The question raised by this correspondence is not whether artificial intelligence belongs in perioperative care. It is who governs it and where its authority ends.

Disclosures

The author declared no potential conflicts of interest with respect to the research, authorship, and publication of this article.

Funding

The author received no financial support for the research, authorship, and publication of this article.

REFERENCES

1. Whittaker G. OpenClaw in 2026: what it is, who's using it, and whether your business should adopt it. *Linux Journal*. April 14, 2026. Accessed July 10, 2026. <https://www.linuxjournal.com/content/openclaw-2026-what-it-whos-using-it-and-whether-your-business-should-adopt-it>
2. Zheng J, Tan J, Lin J. (2026) Understanding and mitigating the risks of OpenClaw for non-technical users: a practical guide with Skill. *arXiv*, arXiv:2606.11007, preprint: not peer reviewed. <https://arxiv.org/abs/2606.11007>
3. n8n for healthcare: secure patient & admin automation. n8n Lab. Accessed July 10, 2026. <https://n8nlab.io/n8n-for-healthcare>